

ARTICLE OPEN



Bacteroidales species in the human gut are a reservoir of antibiotic resistance genes regulated by invertible promoters

Wei Yan¹, A. Brantley Hall^{2,3} and Xiaofang Jiang¹✉

Antibiotic-resistance genes (ARGs) regulated by invertible promoters can mitigate the fitness cost of maintaining ARGs in the absence of antibiotics and could potentially prolong the persistence of ARGs in bacterial populations. However, the origin, prevalence, and distribution of these ARGs regulated by invertible promoters remains poorly understood. Here, we sought to assess the threat posed by ARGs regulated by invertible promoters by systematically searching for ARGs regulated by invertible promoters in the human gut microbiome and examining their origin, prevalence, and distribution. Through metagenomic assembly of 2227 human gut metagenomes and genomic analysis of the Unified Human Gastrointestinal Genome (UHGG) collection, we identified ARGs regulated by invertible promoters and categorized them into three classes based on the invertase-regulating phase variation. In the human gut microbiome, ARGs regulated by invertible promoters are exclusively found in *Bacteroidales* species. Through genomic analysis, we observed that ARGs regulated by invertible promoters have convergently originated from ARG insertions into glycan-synthesis loci that were regulated by invertible promoters at least three times. Moreover, all three classes of invertible promoters regulating ARGs are located within integrative conjugative elements (ICEs). Therefore, horizontal transfer via ICEs could explain the wide taxonomic distribution of ARGs regulated by invertible promoters. Overall, these findings reveal that glycan-synthesis loci regulated by invertible promoters in *Bacteroidales* species are an important hotspot for the emergence of clinically-relevant ARGs regulated by invertible promoters.

npj Biofilms and Microbiomes (2022)8:1; <https://doi.org/10.1038/s41522-021-00260-1>

INTRODUCTION

The proliferation of ARGs has compromised antibiotic treatment for bacteria infections¹. The human gut microbiome is an important reservoir of ARGs^{2–4} and the spread of ARGs from gut microbes to pathogens has been documented⁵. Therefore, ARGs in the human gut microbiome pose a growing threat to human health.

Often, bacteria-carrying ARGs are outcompeted by susceptible strains due to the costs associated with the maintenance and expression of the ARGs^{6–8}. Though it is costly, bacteria can ameliorate the fitness costs of maintaining ARGs through different strategies⁹, such as no-cost, low-cost or gain-of-fitness mutations^{10,11}, compensatory mutations at a second site^{12–14}, or genetic coselection of resistance genes in genetic linkage^{15,16}. Phase-variable antibiotic resistance, which was only recently reported¹⁷, is a newly identified mechanism for antibiotic resistant bacteria to mitigate the fitness cost of encoding ARGs.

Phase variation refers to a reversible change that generates phenotypic variation that helps bacteria adapt to rapidly changing environments^{18,19}. Phase variation often manifests through reversible inversion of DNA regions containing promoters such that in one orientation, a downstream gene is expressed, while in the alternate orientation, the downstream gene is not expressed¹⁷. Such DNA inversions are generally mediated by invertases, which recognize inverted repeats flanking the invertible region and catalyze the reversible inversion^{20–22}. Genes regulated by invertible promoters often contribute to the regulation of characteristics important for bacterial colonization and virulence, including fimbriae^{23,24}, flagella²⁵, and capsular polysaccharides^{26,27}.

Recent advances in computational methods have contributed to the effective identification of the intergenic invertible DNA regions in microbial genomes^{17,28}. ARGs were found to be regulated by

invertible promoters in certain human gut bacteria¹⁷. Bacteria with ARGs regulated by invertible promoters in the ON orientation could be selected for in the presence of antibiotics. Bacteria with ARGs regulated by invertible promoters in the OFF orientation could ameliorate the fitness cost in the absence of antibiotics, which facilitate the maintenance of these ARGs for longer periods in microbial communities¹⁷. The emergence of ARGs regulated by invertible promoters likely increases the burden of combating antibiotic resistance and the spread of these ARGs to pathogens poses an increasing threat to human health. However, many questions that are fundamental to assess the risk of ARGs regulated by invertible promoters remain unanswered, specifically (1) what is the diversity of the invertible promoters regulating ARGs (IP-ARG)? (2) what are the taxonomic boundaries and geographic distribution of the IP-ARG? and (3) how did the IP-ARG originate?

Taking advantage of the considerable amount of human gut metagenomic data generated during the last decade, we aimed to address these questions. Here, we systematically searched for ARGs regulated by invertible promoters through metagenomic assembly of 2227 human gut metagenomes and the Unified Human Gastrointestinal Genome (UHGG) collection of human gut genomes. We found that the taxonomic distribution of ARGs regulated by invertible promoters appears to be restricted to the order *Bacteroidales*. Genomic analysis showed that ARGs regulated by invertible promoters have been convergently derived from ARG insertions into glycan-synthesis loci regulated by invertible promoters. Notably, the identified ARGs regulated by invertible promoters were found to have been mobilized through ICEs and have been widely geographically distributed. Our results reveal the prominent role of glycan-synthesis loci regulated by invertible promoters in *Bacteroidales* species in providing hotspots for the

¹National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA. ²Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland, USA. ³Center for Bioinformatics and Computational Biology, University of Maryland, College Park, Maryland, USA. ✉email: xiaofang.jiang@nih.gov

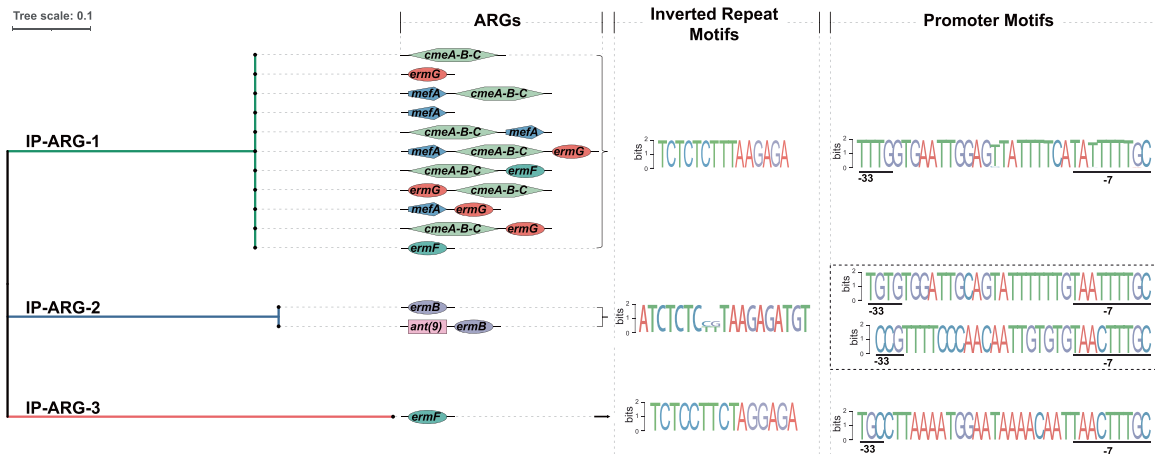


Fig. 1 Three classes of invertible promoters regulating ARGs (IP-ARG) (IP-ARG-1, IP-ARG-2, and IP-ARG-3). The ARGs and putative ARG organization patterns are shown for each class. Different ARGs are shown in different shapes and colors.

emergence of clinically relevant ARGs regulated by invertible promoters.

RESULTS

Identification and classification of invertible promoters regulating ARGs

To expand the known repertoire of ARGs regulated by invertible promoters, we searched publicly available human gut metagenomic datasets, comprising a total of 2227 samples, using the tool PhaseFinder (Supplementary Data 1). Briefly, we assembled the metagenomic data and then searched for regions regulated by invertible promoters with PhaseFinder. PhaseFinder first detects putative invertible regions in the assemblies by identifying inverted repeats, mimics their inversion *in silico*, then checks if there are metagenomic reads supporting both potential orientations to determine whether inversion occurs. We then identified putative invertible promoters in the regions by searching for promoter sequences in the identified invertible regions. We then scanned the operons downstream of the invertible promoters to find those containing ARGs. This search uncovered 62 contigs, from samples collected from 47 individuals, containing invertible promoters regulating ARGs (IP-ARG). Redundant instances in the same individual were removed. In total, 48 IP-ARG instances were identified. All identified IP-ARG were located immediately downstream of a gene encoding an invertase and have promoters with an identical or close match to the consensus-promoter sequence -33 (TTTG) and -7 (TANNTTGY)²⁹.

We constructed a phylogenetic tree based on the nucleotide sequence of the invertases. We found that the invertible promoters regulating ARGs can be grouped based on the invertase into three distinct classes, denoted IP-ARG-1, IP-ARG-2, and IP-ARG-3 (Fig. 1, Supplementary Data 2). The nucleotide sequences of the invertase of the same class are nearly identical. Of the three classes, class IP-ARG-1 was found to contain the most identified instances (42 out of 48).

To further characterize these three classes of IP-ARG, we analyzed the motifs of inverted repeats and invertible promoters as well as the ARGs regulated by the invertible regions. The invertible promoter motifs varied among classes but were nearly identical within the same class (Fig. 1). The inverted repeat motifs were found to be similar across different classes. Interestingly, we found two invertible regions located immediately downstream of the invertase gene in class IP-ARG-2 (Fig. 1). The two invertible regions were located adjacent to each other and the inverted repeats were similar in these two regions. This suggested that these two invertible regions could be regulated by the same upstream

invertase gene. However, the invertible promoters were different from each other (Fig. 1, Supplementary Data 2). This suggested that there may be differences in the strength of the promoters leading to differences in expression level. The ARGs regulated by invertible promoters varied among classes (Fig. 1, Supplementary Data 2). Most invertible promoters of class IP-ARG-1 regulated the *cmeABC* operon, *ermG* gene, or both, while some also regulated other ARGs, including *mefA*, *ermF*, or *tetQ*. These ARGs could confer resistance to diverse antibiotics, such as fluoroquinolone, macrolides, lincosamides, streptogramin, and tetracycline^{30–34}. The *ermB* and *ermF* genes, which confer resistance to streptogramin, macrolides, and lincosamides^{31,33}, were found to be regulated by class IP-ARG-2 and class IP-ARG-3, respectively. In addition, class IP-ARG-2 was also found to regulate an *ant(9)* gene homolog which confers resistance to aminoglycosides^{35,36}.

We analyzed the promoter orientation in the metagenomic samples and found that in only 4 out of 96 cases (with total supporting reads >20), the promoters were predominantly in the ON orientation while the remainder were in the OFF orientation (Supplementary Data 3). Of the four cases, three have been reported previously¹⁷ and promoters in the ON orientation have been found to be selected for by antibiotics, while the fourth has no metadata available in terms of antibiotic treatment. This supports that in all the four types of invertible promoters, the OFF orientation is advantageous in most samples, possibly by mitigating the fitness cost of expressing ARGs in the absence of antibiotics.

Taxonomic distribution of IP-ARG

We expanded the search to include genomes from the Unified Human Gastrointestinal Genome (UHGG) collection³⁷ by detecting nearly identical (>99% identity) invertase genes that are immediately upstream of invertible promoters regulating ARGs. These invertase genes were grouped into the corresponding classes of IP-ARG (Supplementary Data 4). Consistent with the finding from metagenomic contigs, most ARGs regulated by invertible promoters belonged to class IP-ARG-1 (176 out of 210 genomes). Based on the metadata from the metagenomic samples used in this study and provided by the UHGG collection, we found that ARGs regulated by invertible promoters were observed in multiple countries, but the geographic prevalence varied across classes (Table 1, Supplementary Data 2 and 4). Examples of IP-ARG-1 and IP-ARG-2 were identified in metagenomes from 17 and 7 countries, respectively, spanning three continents (Asia, Europe, and North America). Class IP-ARG-3 were only observed in the metagenomes from Denmark, which might be due to the limited sampling of publicly-available metagenomic data. The results reveal that

Table 1. Taxonomic and geographic distribution of ARGs regulated by invertible promoters.

IP-ARG group ^a	Order ^b	Family	Genus	Species	Country	
1	<i>Bacteroidales</i>	<i>Bacteroidaceae</i>	<i>Bacteroides</i>	<i>B. cellulosilyticus</i> , <i>B. eggerthii</i> ,	Austria, Canada, China, Denmark, Estonia, Finland, France, Germany, Israel, Japan, Kazakhstan, Netherlands, Russia, Spain, Sweden, United Kingdom, United States	
				<i>B. fluxus</i> , <i>B. fragilis</i> ,		
				<i>B. intestinalis</i> , <i>B. ovatus</i> ,		
				<i>B. stercoris</i> , <i>B. thetaiotaomicron</i> ,		
				<i>B. uniformis</i>		
				<i>Phocaeicola</i>		<i>P. coprocola</i> , <i>P. coprophilus</i> ,
				<i>P. dorei</i> , <i>P. sp000432735</i> ,		
				<i>P. vulgatus</i>		
				<i>Prevotella</i>		<i>P. copri</i> , <i>P. sp000834015</i> ,
				<i>P. sp001275135</i> ,		
<i>P. sp900313215</i> , <i>P. stercorea</i>						
	<i>Barnesiellaceae</i>	<i>Barnesiella</i>	<i>B. intestinihominis</i>			
	<i>Coprobacteraceae</i>	<i>Coprobacter</i>	<i>C. fastidiosus</i>			
	<i>Tannerellaceae</i>	<i>Parabacteroides</i>	<i>P. distasonis</i> , <i>P. goldsteinii</i> , <i>P. merdae</i>			
2	<i>Bacteroidales</i>	<i>Bacteroidaceae</i>	<i>Bacteroides</i>	<i>B. caccae</i> , <i>B. eggerthii</i> ,	China, Denmark, Estonia, Israel, Italy, United Kingdom, United States	
				<i>B. fragilis</i> ,		
				<i>B. thetaiotaomicron</i> ,		
				<i>B. uniformis</i>		
				<i>Paraprevotella</i>		<i>P. xylaniphila</i>
	<i>Phocaeicola</i>	<i>P. dorei</i>				
3	<i>Bacteroidales</i>	<i>Bacteroidaceae</i>	<i>Phocaeicola</i>	<i>P. dorei</i>	Denmark	
		UBA932	RC9	RC9 sp000434935		

^aThe class of invertible promoters regulating ARGs (IP-ARG).

^bTaxonomic information of the bacteria host at order, family, genus, and species levels. The taxonomic information is annotated based on GTDB release95.

species harboring ARGs regulated by invertible promoters are widely geographically distributed.

IP-ARG was found to be exclusively distributed *Bacteroidales* species (Table 1, Supplementary Data 2 and 4). Class IP-ARG-1 was observed in 24 *Bacteroidales* species from the families *Bacteroidaceae*, *Barnesiellaceae*, *Coprobacteraceae*, and *Tannerellaceae*. Class IP-ARG-2 could be identified in 7 species from *Bacteroidaceae* that belonged to the order *Bacteroidales*. Class IP-ARG-3 was found in two species, *Phocaeicola dorei* and a novel species RC9 sp000434935, which also belongs to the *Bacteroidales* order. The wide taxonomic distribution combined with the sparse occurrence of these ARGs and the presence of nearly identical invertases in a broad range of *Bacteroidales* hosts is not consistent with vertical transmission, suggesting that ARGs regulated by invertible promoters could be horizontally transferred by mobile genetic elements (MGEs).

All three classes of IP-ARG are located within integrative conjugative elements

To determine whether IP-ARG was on MGEs, we searched the invertases and invertible regions regulating ARGs as well as the flanking sequences against ImmeDB³⁸ and the ICEberg database³⁹. We found that all three classes of IP-ARG are located within integrative conjugative elements (ICEs) (Fig. 2, Supplementary Fig. 1, and Supplementary Data 4). The majority of examples from class IP-ARG-1 were identified within ICEs related to ICE26 in ImmeDB (Fig. 2a). ARGs regulated by invertible promoters from class IP-ARG-1 were also found within another novel ICE (Fig. 2a). This suggests that the progenitor of IP-ARG-1 might have emerged

before subsequently inserting into multiple ICEs. Examples of class IP-ARG-2 were detected in ICEs related to ICE14 in ImmeDB (Fig. 2b). Class IP-ARG-3 that is contained within an ~50-kb sequence fragment most closely matches ICE34 in ImmeDB (Fig. 2c). The fact that all three classes of IP-ARG are located within ICEs may allow for further dissemination in the human population because ICEs encode the necessary machinery to horizontally transfer between species mobilizing any ARGs that they acquire.

ARGs regulated by invertible promoters were frequently observed to be included in a highly variable region that was located immediately downstream of the invertible regions. The highly varied regions contained genes that were not necessary for the ICE replication and transfer, but often important for conferring selective and adaptive advantages for hosts in the changing environments (Fig. 2). Most of the variable genes in the highly varied region were ARGs in ICEs containing class IP-ARG-1 (Fig. 2a), which suggested multiple insertions of ARGs at this region. In ICEs carrying class IP-ARG-2, not only the ARGs but also the glycan-synthesis genes, even clusters that contained only glycan-synthesis genes, were located in the regions downstream of the invertases (Fig. 2b). In different ICEs with class IP-ARG-3, the highly variable regions contained varied genes or operons, such as ARGs, operons involved in pathogenicity island or T6SS, or integrative and mobilizable elements encoding ARG *tetQ* (Fig. 2c). All these indicated that the loci in such highly variable regions downstream of the invertible promoters were hotspots for the acquisition of elements, including ARGs. The newly acquired ARGs can then be subsequently horizontally transferred to a wide range of *Bacteroidales* species, serving as a reservoir.

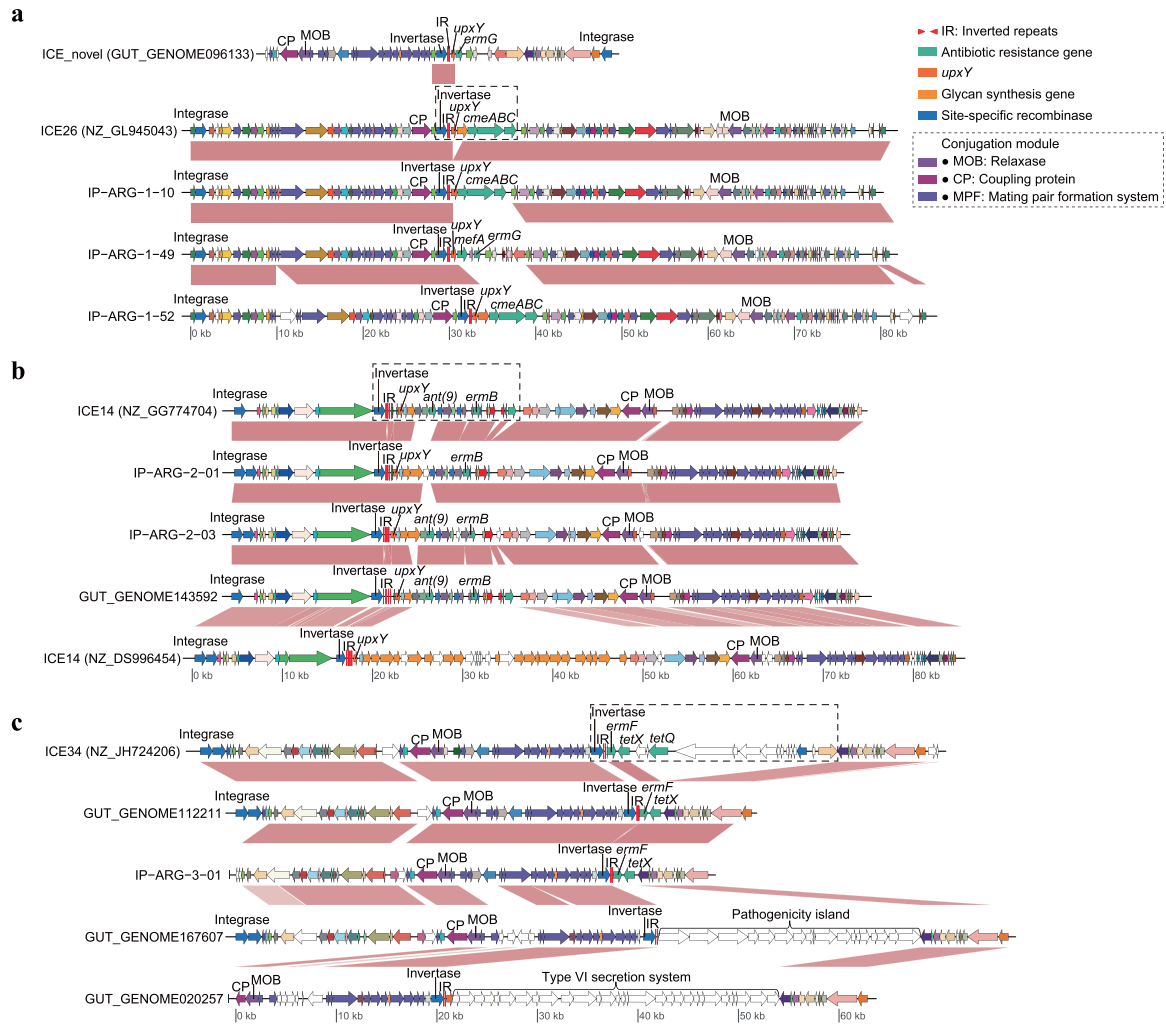


Fig. 2 Genomic comparison and context analysis of representative contigs carrying IP-ARG-1 (a), IP-ARG-2 (b), and IP-ARG-3 (c) as well as related elements. The regions located adjacent to the invertible promoters (black-dotted boxes) were found to be highly variable across element variants within each ICE. ICE26, ICE14, and ICE34 are ICE accession numbers from ImmeDB. ICE_novel1 is a newly identified ICE that has not been included in the ImmeDB database. The NCBI genome-accession numbers are shown in the parentheses after the ICE accession numbers. The sequence labels that start with GUT_GENOME are genome-accession numbers of the UHGG database. Orthologous genes are plotted with the same color and are linked by pink connections. Site-specific recombinase, antibiotic resistance, and glycan-synthesis genes are colored blue, light green, and orange, respectively. The genes that do not have orthologs are white.

The invertible regions regulating ARGs appear to originate from those regulating glycan-synthesis genes independently and convergently

To better understand the evolution of IP-ARG, we performed comparative analyses on the highly varied regions regulated by invertible promoters. We found that glycan-synthesis genes, such as *wecA* gene, or genes involved in synthesis or regulation of capsular polysaccharide synthesis, such as the *upxY* gene⁴⁰, were frequently located immediately downstream of the invertible promoters and upstream of the identified ARGs (Figs. 2a, b and 3). In addition, the inverted repeats of class IP-ARG-2 in the genome GUT_GENOME143592 are identical to the inverted repeats of the invertible promoter regulating glycan synthesis cluster on ICE14 (NZ_DS996454). The invertase amino acid sequences of these two are more than 96% identical based on BLAST analysis⁴¹. These results support that the invertible regions regulating ARGs share a common evolutionary origin with invertible regions regulating glycan-synthesis genes.

Gene-context analysis further shows that invertible regions regulating ARGs appeared to be derived from invertible regions originally regulating glycan-synthesis genes. ARGs regulated by

invertible promoters were frequently found to be located mostly downstream or in a few cases upstream of glycan-synthesis genes (Figs. 2a, b and 3). The *upxY* gene, a gene that is often the first gene in polysaccharide-biosynthesis operons, is also the first gene in the several ARG operons regulated by invertible promoters (Fig. 3). BLASTn comparisons between the highly variable regions carrying IP-ARG-1, showed that it is likely that *upxY* is completely degenerated in IP-ARG-1-14 and IP-ARG-1-15, partially degenerated in IP-ARG-1-39, and fused with the partial sequence of the *wecA* gene in IP-ARG-1-01 (Fig. 3).

To understand the evolutionary history of different classes of IP-ARG, a phylogenetic tree was constructed based on the protein sequences of the local invertase genes (Fig. 4). We rooted the invertase phylogenetic tree with *Bacteroidales* tyrosine integrases as the outgroup and found that the invertases of all three classes of IP-ARG are located in a clade where invertases were frequently mobilized by conjugative elements. Invertases of class IP-ARG-1 were found in two different ICEs, likely indicating that the ARG was inserted adjacent to the invertase before they were transferred together to another ICE. The closest relative of class IP-ARG-2 is an invertase of glycan-synthesis locus, which is also located on the ICE of

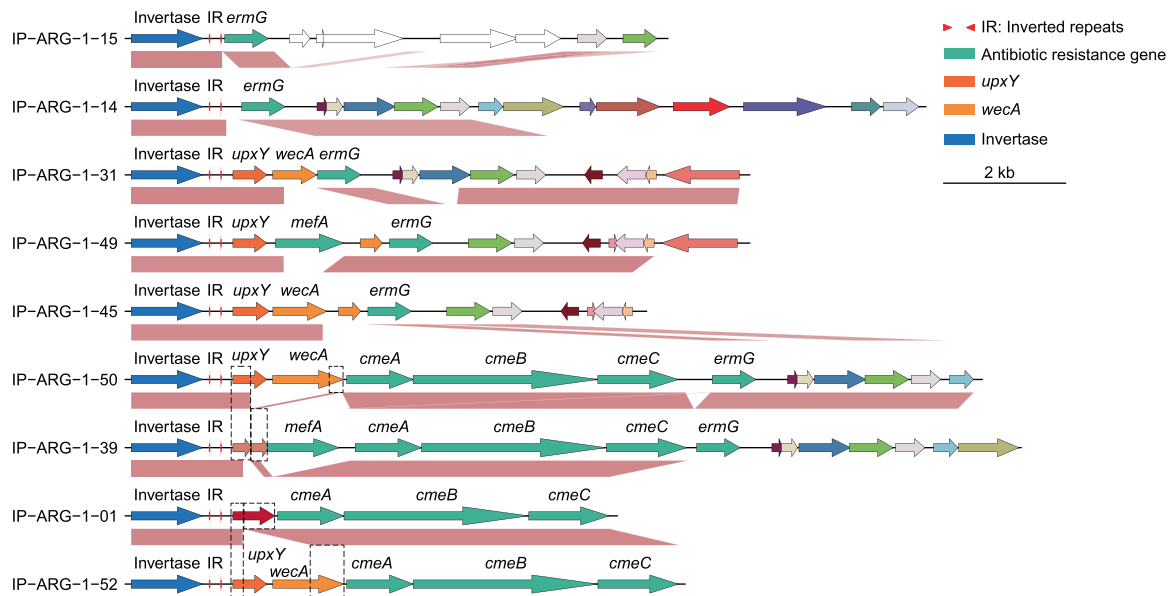


Fig. 3 Comparisons of highly variable regions carrying IP-ARG-1 demonstrate the degeneration of the *upxY* gene. Different degeneration statuses of the *upxY* gene are: completely degenerated (IP-ARG-1-14 and IP-ARG-1-15), partially degenerated (IP-ARG-1-39), fusion gene (IP-ARG-1-01), and intact gene (the remainder). Orthologous genes are plotted with the same color and are linked by pink connections. Site-specific recombinase, antibiotic resistance, and glycan-synthesis genes are colored blue, light green, and orange, respectively. The genes that do not have orthologs are white.

the ICE14 family. Given the fact that neighboring branches are mostly glycan-synthesis loci, the most parsimonious explanation is that class IP-ARG-2 emerged as the ARG inserted into the glycan-synthesis operon localized in ICEs. Class IP-ARG-3 branches early in the tree, suggesting that the emergence of class IP-ARG-3 was likely independent of class IP-ARG-1 or IP-ARG-2. Hence, the emergence of different classes of IP-ARG might be the result of convergent evolution and the evolutionary events that led to the emergence of ARGs regulated by invertible promoters likely occurred independently at least three times. Given the large variety of glycan-synthesis loci regulated by invertible promoters in *Bacteroidales* species, this finding suggests that new classes of IP-ARG might readily emerge.

DISCUSSION

Our study is a comprehensive investigation of the origin, evolution, and prevalence of invertible promoters regulating ARGs. Our analysis revealed that ARGs regulated by invertible promoters are (a) exclusively found in *Bacteroidales* species, (b) often originate from ARG insertions into glycan-synthesis loci regulated by invertible promoters, (c) frequently mobilized by ICEs that may explain their wide taxonomic distribution within the *Bacteroidales* and their rapid dissemination, and (d) widely geographically distributed.

The IP-ARG found in this study was based on *in silico* prediction and further experiments are needed to confirm if these ARGs are indeed subject to phase variation. In the previous study, the phase-variable expression of ARGs regulated by invertible promoters was experimentally verified¹⁷. In this study, several of the identified ARGs are farther from the promoter, and in some cases, the *upxY* gene downstream of the invertible promoter region is degenerated. Due to its role in transcriptional antitermination, degeneration of *upxY* may lead to ARGs regulated by invertible promoters not being expressed⁴⁰. Our analysis only identified ARGs regulated by invertible promoters in *Bacteroidales* species. There could be several reasons for this observed taxonomic restriction. First, gut *Bacteroidales* genomes typically contain numerous loci regulated by invertible promoters. In some *Bacteroides* species, such as *Bacteroides fragilis*, there are more than 20 such loci including up to 7 capsular polysaccharide loci^{42,43}.

While other taxa prevalent in the gut, including Proteobacteria, Firmicutes, and Actinobacteria, have loci regulated by invertible promoters, these loci are far less common and have far fewer examples per genome^{17,44}. Therefore, ARG insertions into *Bacteroidales* loci regulated by invertible promoters are more probable than insertions into loci regulated by invertible promoters in other phyla. Furthermore, as abundant members of the human gut microbiome, metagenomic sequencing leads to high coverage of *Bacteroidales* species, which increases the likelihood that a region regulated by invertible promoters can be detected with PhaseFinder¹⁷. Though more than two thousand metagenomic samples were analyzed, no ARGs regulated by invertible promoters were identified in other taxa. This might be due to the limitation of PhaseFinder or suggests that ARGs regulated by invertible promoters might be less common in taxa other than *Bacteroidales*.

This finding highlights a potential threat to human health that *Bacteroidales* species pose as a reservoir for the dissemination of ARGs regulated by invertible promoters. *Bacteroidales* is one of the most abundant taxa in the human gut and *Bacteroidales* species are regarded as a reservoir of ARGs⁴⁵. Moreover, members of *Bacteroidales* species, such as *Bacteroides fragilis*, are considered opportunistic pathogens⁴⁶ and can be the causative agent of appendicitis and intra-abdominal abscesses^{47,48}. Antibiotics have been used to treat such infections, but an increasing rate of antibiotic resistance has been noted in *Bacteroidales* species⁴⁹⁻⁵². ARGs regulated by invertible promoters may contribute to the continued maintenance of clinically relevant ARGs such as *cmeABC*, *ermF*, and *tetQ*, which confer resistance to widely used antibiotics including macrolides, streptogramin and tetracycline^{30,33,34}. Most species identified with ARGs regulated by invertible promoters are considered symbionts in the human gut. However, ARG transfer from symbionts to pathogens through horizontal transfer has been documented^{53,54}. The transfer of ARGs regulated by invertible promoters from symbionts to pathogens via MGEs, such as the ICEs identified in this study, might promote resistance to a wide array of antibiotics in pathogenic species⁵⁵, posing a threat to public health in the future. Due to the fact that the transfer of ICEs from *Bacteroidales* to other orders is rare³⁸, the spread of ARGs regulated by

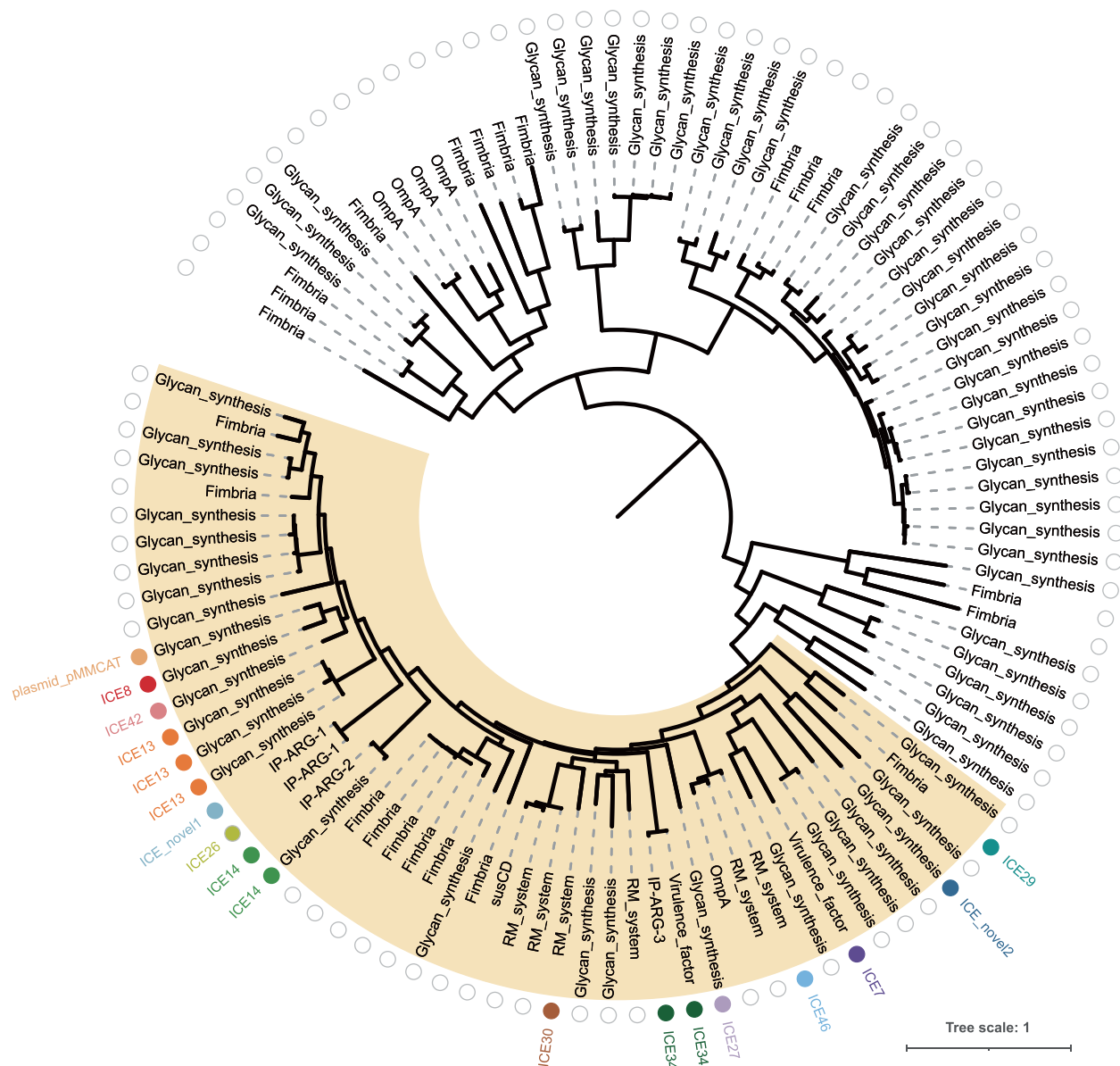


Fig. 4 The evolutionary events that led to the emergence of ARGs regulated by invertible promoters may have occurred independently at least three times as the result of convergent evolution. The phylogenetic tree is inferred based on the alignment of protein sequences of invertases. The invertase tree is rooted with *Bacteroidales* tyrosine integrases as the outgroup. Invertases are labeled with the functional annotation of the loci regulated by the invertase and invertible promoters. The empty gray circle indicates that the invertase is not on a conjugative element, while the solid-colored circle indicates that the invertase is on a conjugative element. Different conjugative elements are distinguished by different colors. The plasmid in the tree is a conjugative plasmid pMMCAT⁴⁹. The ICE names are based on the ImmeDB database and the ICE_novelt1 and 2 are newly identified ICEs that have not been included in the ImmeDB database. The clade where invertases were frequently mobilized by conjugative elements is colored light yellow.

invertible promoters is likely to be limited within the order *Bacteroidales*. As such, monitoring these *Bacteroidales* species, especially the ICEs in these species, might be important in mitigating the threat of ARGs regulated by invertible promoters to human health.

METHODS

De novo assembly and gene annotation of metagenomic datasets

We downloaded metagenomic sequencing data that consisted of 2227 human gut samples encompassing seven studies (Supplementary Data 1). Low-quality reads were removed and sequencing adapters were trimmed with trim_galore (v0.6.4, <https://github.com/FelixKrueger/TrimGalore>). The

filtered data were mapped to the human genome (hg19) using bowtie2 (v2.3.5.1)⁵⁶ to filter human reads. The cleaned reads from each sample were assembled with SPAdes (v3.14.0) using the --meta option^{57,58}, and sequences less than 500 bp were removed. Gene annotation was performed using Prokka (v1.14.5)⁵⁹ with the parameter --metagenome.

Identification of putative invertible promoters

PhaseFinder (v1.0)¹⁷ was employed to identify putative invertible regions in the metagenomic assemblies. The default parameters of PhaseFinder were used. We filtered the results by removing the invertible DNA regions with <2 reads supporting the R orientation from the paired-end method, and the Pe_ratio <1%. Furthermore, the invertible DNA regions containing or overlapping coding sequences (CDS) were removed. We used FIMO⁶⁰ to search for promoter sequences in invertible regions based on previously

identified promoter motifs¹⁷. Only invertible regions with promoters detected were retained for further analysis. The sequence-motif logos of aligned promoters and aligned inverted repeats were generated with WebLogo (version 2.8.2)⁶¹.

Annotation of ARGs

We extracted the genes in the operons downstream the putative invertible promoters. We searched the protein sequences of these genes for known ARGs from the Comprehensive Antibiotic Resistance Database (CARD) (v3.0.7)⁶² using BLASTn (v2.10.0)⁴¹. The BLAST results were filtered using the parameters: `-perc_identity 80, -evaluate 1e-10, and -culling_limit 1`. Resistance Gene Identifier (RGI, v5.0.0)⁶² was also used to predict known antibiotic-resistance elements using the following parameters: `rgi main -t contig -a BLAST -n 8 -d wgs --local`. ARGs located in the operon immediately downstream of the identified invertible promoters were kept for further analysis. Pairwise BLASTn searches were performed between each pair of contigs with putative IP-ARG from samples of the same individual. If multiple contigs from samples of the same individual likely originated from the same genomic region, only the longest contig was selected as a representative in further analysis.

Identification of host species and mobile genetic elements

The identified contigs containing ARGs regulated by invertible promoters were searched with BLASTn against the NCBI nonredundant nucleotide (nt) database and the Unified Human Gastrointestinal Genome UHGG database³⁷ with an e-value threshold $<1e-10$. If all best BLAST hits for a contig were from the same species, the species taxonomy was assigned to the contig. If there are ambiguous matches (hits from multiple species with the same top-match statistics), the lowest common ancestor of all the hits was assigned as the taxonomy of the contig. Next, we performed BLASTn to search for identical homologs ($>99\%$ identity) of the identified invertases and invertible regions against the 204,938 nonredundant genomes from the UHGG database. The identical homologs were grouped into corresponding classes of IP-ARG and the UHGG annotations for host species were examined to identify host species for IP-ARG.

To identify integrative and conjugative elements (ICEs) in different classes of IP-ARG, we searched the ICEberg^{39,63} and ImmeDB³⁸ databases using BLASTn with an e-value $<1e-10$. We used ConjScan^{64,65} via a Galaxy web server (<https://galaxy.pasteur.fr>) to annotate genes involved in conjugation modules in ICEs.

Genomic comparison and phylogenetic analysis

The bacteria genomes and genome comparisons were visualized with the R package genoPlotR. Genes were colored based on their predicted function. The function assignment of genes was first based on the prokka annotation. If a gene was annotated as “hypothetical protein” by prokka, we then performed BLAST search on the protein sequences of this gene against NCBI nr database as well as a small curated database of genes reported to be involved in polysaccharide biosynthesis⁴². The best hits were used to assign function to the genes.

To understand the evolutionary history of different classes of IP-ARG, a phylogenetic tree was constructed based on the invertases. The protein sequences of the invertase regulating ARGs regulated by invertible promoters identified in this study and those that have been previously reported¹⁷ were extracted. We also added invertases identified in the metagenomic assemblies based on PhaseFinder to the invertases dataset. If the invertase is located on a contig that lacks genomic context to determine if it is located on a mobile genetic element, the invertase was removed from further analysis. Redundant invertase genes were filtered using cd-hit with the 99% identity threshold⁶⁶. Multiple alignments of protein sequences of the invertases were performed with MUSCLE (v3.8.31)⁶⁷. The alignment results were analyzed in FastTree (v2.1.10)⁶⁸ with default parameters to infer the phylogenetic trees. Only the invertase genes regulating functionally characterized genes or operons were included in the tree. *Bacteroidales* integrases³⁸ were added to the alignments and used as the outgroup to root the tree. The phylogenetic tree was visualized using iTOL (<https://itol.embl.de>).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The human gut metagenomic data used in this study are downloaded from the NCBI. Accession numbers are listed in Supplementary Data 1. Metagenomic assemblies are available on the NCBI FTP site at https://ftp.ncbi.nlm.nih.gov/pub/mgx/IP-ARG_assemblies/. Nucleotide sequences of the identified contigs carrying ARGs regulated by invertible promoters can be accessed at https://ftp.ncbi.nlm.nih.gov/pub/mgx/IP-ARG_assemblies/IPARG_contigs.fasta.

Received: 24 March 2021; Accepted: 26 November 2021;

Published online: 10 January 2022

REFERENCES

- Alekshun, M. N. & Levy, S. B. Molecular mechanisms of antibacterial multidrug resistance. *Cell* **128**, 1037–1050 (2007).
- Sommer, M. O. A., Dantas, G. & Church, G. M. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128–1131 (2009).
- Kim, S., Covington, A. & Pamer, E. G. The intestinal microbiota: antibiotics, colonization resistance, and enteric pathogens. *Immunol. Rev.* **279**, 90–105 (2017).
- Wallace, M. J., Fishbein, S. R. S. & Dantas, G. Antimicrobial resistance in enteric bacteria: current state and next-generation solutions. *Gut Microbes* **12**, 1799654 (2020).
- Broaders, E., Gahan, C. G. M. & Marchesi, J. R. Mobile genetic elements of the human gastrointestinal tract: potential for spread of antibiotic resistance genes. *Gut Microbes* **4**, 271–280 (2013).
- Bilgin, N., Claessens, F., Pahverk, H. & Ehrenberg, M. Kinetic properties of *Escherichia coli* ribosomes with altered forms of S12. *J. Mol. Biol.* **224**, 1011–1027 (1992).
- Andersson, D. I. & Levin, B. R. The biological cost of antibiotic resistance. *Curr. Opin. Microbiol.* **2**, 489–493 (1999).
- Hernando-Amado, S., Sanz-García, F., Blanco, P. & Martínez, J. L. Fitness costs associated with the acquisition of antibiotic resistance. *Essays Biochem* **61**, 37–48 (2017).
- Andersson, D. I. & Hughes, D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat. Rev. Microbiol.* **8**, 260–271 (2010).
- Sander, P. et al. Fitness cost of chromosomal drug resistance-conferring mutations. *Antimicrobial Agents Chemother.* **46**, 1204–1211 (2002).
- Ramadhan, A. A. Survivability of vancomycin resistant enterococci and fitness cost of vancomycin resistance acquisition. *J. Clin. Pathol.* **58**, 744–746 (2005).
- Schrag, S. J. & Perrot, V. Reducing antibiotic resistance. *Nature* **381**, 120–121 (1996).
- Björkman, J., Nagaev, I., Berg, O. G., Hughes, D. & Andersson, D. I. Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. *Science* **287**, 1479–1482 (2000).
- Comas, I. et al. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat. Genet.* **44**, 106–110 (2011).
- Enne, V. I., Bennett, P. M., Livermore, D. M. & Hall, L. M. C. Enhancement of host fitness by the *sul2*-coding plasmid p9123 in the absence of selective pressure. *J. Antimicrob. Chemother.* **53**, 958–963 (2004).
- Roberts, A. P. & Mullany, P. Tn916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiol. Rev.* **35**, 856–871 (2011).
- Jiang, X. et al. Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut. *Science* **363**, 181–187 (2019).
- Reyes Ruiz, L. M., Williams, C. L. & Tamayo, R. Enhancing bacterial survival through phenotypic heterogeneity. *PLoS Pathog.* **16**, e1008439 (2020).
- Trzilova, D. & Tamayo, R. Site-specific recombination - how simple DNA inversions produce complex phenotypic heterogeneity in bacterial populations. *Trends Genet* **37**, 59–72 (2021).
- Henderson, I. R., Owen, P. & Nataro, J. P. Molecular switches—the ON and OFF of bacterial phase variation. *Mol. Microbiol.* **33**, 919–932 (1999).
- van der Woude, M. W. Phase variation: how to create and coordinate population diversity. *Curr. Opin. Microbiol.* **14**, 205–211 (2011).
- Villemur, R. & Déziel, E. Phase variation and antigenic variation. *Dyn. Bacterial Genome*. <https://doi.org/10.1017/cbo9780511541544.008> (2005).
- Abraham, J. M., Freitag, C. S., Clements, J. R. & Eisenstein, B. I. An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **82**, 5724–5727 (1985).
- Olsen, P. B. & Klemm, P. Localization of promoters in the *fim* gene cluster and the effect of H-NS on the transcription of *fimB* and *fimE*. *FEMS Microbiol. Lett.* **116**, 95–100 (1994).

25. Bonifield, H. R. & Hughes, K. T. Flagellar phase variation in *Salmonella enterica* is mediated by a posttranscriptional control mechanism. *J. Bacteriol.* **185**, 3567–3574 (2003).
26. Coyne, M. J., Chatzidaki-Livanis, M., Paoletti, L. C. & Comstock, L. E. Role of glycan synthesis in colonization of the mammalian gut by the bacterial symbiont *Bacteroides fragilis*. *Proc. Natl Acad. Sci. USA* **105**, 13099–13104 (2008).
27. Krinos, C. M. et al. Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature* **414**, 555–558 (2001).
28. Sekulovic, O. et al. Genome-wide detection of conservative site-specific recombination in bacteria. *PLoS Genet.* **14**, e1007332 (2018).
29. Bayley, D. P., Rocha, E. R. & Smith, C. J. Analysis of *cepA* and other *Bacteroides fragilis* genes reveals a unique promoter structure. *FEMS Microbiol. Lett.* **193**, 149–154 (2000).
30. Pumbwe, L. et al. *Bacteroides fragilis* BmeABC efflux systems additively confer intrinsic antimicrobial resistance. *J. Antimicrob. Chemother.* **58**, 37–46 (2006).
31. Shoemaker, N. B., Vlamakis, H., Hayes, K. & Salyers, A. A. Evidence for extensive resistance gene transfer among *Bacteroides* spp. and among *Bacteroides* and other genera in the human colon. *Appl. Environ. Microbiol.* **67**, 561–568 (2001).
32. Wang, Y., Wang, G.-R., Shelby, A., Shoemaker, N. B. & Salyers, A. A. A newly discovered *Bacteroides* conjugative transposon, CTnGERM1, contains genes also found in gram-positive bacteria. *Appl. Environ. Microbiol.* **69**, 4595–4603 (2003).
33. Halula, M. C., Manning, S. & Macrina, F. L. Nucleotide sequence of *ermFU*, a macrolide-lincosamide-streptogramin (MLS) resistance gene encoding an RNA methylase from the conjugal element of *Bacteroides fragilis* V503. *Nucleic Acids Res.* **19**, 3453 (1991).
34. Bartha, N. A., Sóki, J., Urbán, E. & Nagy, E. Investigation of the prevalence of *tetQ*, *tetX* and *tetX1* genes in *Bacteroides* strains with elevated tetracycline minimum inhibitory concentrations. *Int. J. Antimicrob. Agents* **38**, 522–525 (2011).
35. Bryan, L. E., Kowand, S. K. & Van Den Elzen, H. M. Mechanism of aminoglycoside antibiotic resistance in anaerobic bacteria: clostridium perfringens and *Bacteroides fragilis*. *Antimicrob. Agents Chemother.* **15**, 7–13 (1979).
36. Murphy, E. Nucleotide sequence of a spectinomycin adenyltransferase AAD(9) determinant from *Staphylococcus aureus* and its relationship to AAD(3") (9). *Mol. Gen. Genet.* **200**, 33–39 (1985).
37. Almeida, G. et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-020-0603-3> (2020).
38. Jiang, X., Hall, A. B., Xavier, R. J. & Alm, E. J. Comprehensive analysis of chromosomal mobile genetic elements in the gut microbiome reveals phylum-level niche-adaptive gene pools. *PLoS One* **14**, e0223680 (2019).
39. Liu, M. et al. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res* **47**, D660–D665 (2019).
40. Chatzidaki-Livanis, M., Coyne, M. J. & Comstock, L. E. A family of transcriptional antitermination factors necessary for synthesis of the capsular polysaccharides of *Bacteroides fragilis*. *J. Bacteriol.* **191**, 7288–7295 (2009).
41. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
42. Coyne, M. J. & Comstock, L. E. Niche-specific features of the intestinal bacteroidales. *J. Bacteriol.* **190**, 736–742 (2008).
43. Patrick, S. et al. Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of *Bacteroides fragilis*. *Microbiology* **156**, 3255–3269 (2010).
44. van der Woude, M. W. & Bäuml, A. J. Phase and antigenic variation in bacteria. *Clin. Microbiol. Rev.* **17**, 581–611 (2004).
45. Salyers, A., Gupta, A. & Wang, Y. Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends Microbiol.* **12**, 412–416 (2004).
46. Wexler, H. M. *Bacteroides*: the good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* **20**, 593–621 (2007).
47. Bennion, R. S. et al. The bacteriology of gangrenous and perforated appendicitis—revisited. *Ann. Surg.* **211**, 165–171 (1990).
48. Jotwani, R., Gupta, U., Watanabe, K. & Ueno, K. Pathogenicity of *Bacteroides fragilis* group in rat intra-abdominal abscesses. *Microbiol. Immunol.* **36**, 1041–1049 (1992).
49. Snyderman, D. R. et al. Lessons learned from the anaerobe survey: historical perspective and review of the most recent data (2005–2007). *Clin. Infect. Dis.* **50**, S26–S33 (2010).
50. Nagy, E., Urbán, E. & Nord, C. E. ESCMID Study Group on Antimicrobial Resistance in Anaerobic Bacteria. Antimicrobial susceptibility of *Bacteroides fragilis* group isolates in Europe: 20 years of experience. *Clin. Microbiol. Infect.* **17**, 371–379 (2011).
51. Eitel, Z., Sóki, J., Urbán, E. & Nagy, E., ESCMID Study Group on Anaerobic Infection. The prevalence of antibiotic resistance genes in *Bacteroides fragilis* group strains isolated in different European countries. *Anaerobe* **21**, 43–49 (2013).
52. Veloo, A. C. M., Baas, W. H., Haan, F. J., Coco, J. & Rossen, J. W. Prevalence of antimicrobial resistance genes in *Bacteroides* spp. and *Prevotella* spp. Dutch clinical isolates. *Clin. Microbiol. Infect.* **25**, 1156.e9–1156.e13 (2019).
53. Stecher, B. et al. Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proc. Natl Acad. Sci. USA* **109**, 1269–1274 (2012).
54. Thanh Duy, P. et al. Commensal *Escherichia coli* are a reservoir for the transfer of XDR plasmids into epidemic fluoroquinolone-resistant *Shigella sonnei*. *Nat. Microbiol.* **5**, 256–264 (2020).
55. Botelho, J. & Schulenburg, H. The role of integrative and conjugative elements in antibiotic resistance evolution. *Trends Microbiol.* <https://doi.org/10.1016/j.tim.2020.05.011> (2020)..
56. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
57. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
58. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
59. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
60. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
61. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
62. Jia, B. et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).
63. Bi, D. et al. ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. *Nucleic Acids Res.* **40**, D621–D626 (2012).
64. Abby, S. S., Néron, B., Ménager, H., Touchon, M. & Rocha, E. P. C. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS ONE* **9**, e110726 (2014).
65. Abby, S. S. et al. Identification of protein secretion systems in bacterial genomes. *Sci. Rep.* **6**, 1–14 (2016).
66. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
67. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinforma.* **5**, 113 (2004).
68. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
69. García-Bayona, L., Coyne, M. J. & Comstock, L. E. Mobile Type VI secretion system loci of the gut Bacteroidales display extensive intra-ecosystem transfer, multi-species spread and geographical clustering. *PLoS Genet.* **17**, e1009541 (2021).

ACKNOWLEDGEMENTS

W.Y. and X.J. are supported by the Intramural Research Program of the NIH, National Library of Medicine. B.H. is supported by startup funding from the University of Maryland. We thank NIH's Biowulf cluster team. This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>).

AUTHOR CONTRIBUTIONS

X.J. conceived and designed the study. W.Y. performed the data analysis and wrote the draft paper. X.J. and A.B.H. reviewed and revised the paper. X.J. supervised the work. All authors read and approved the final paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41522-021-00260-1>.

Correspondence and requests for materials should be addressed to Xiaofang Jiang.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2022